Do humans spontaneously take the perspective of others?

Geoff G. Cole[1], Mark Atkinson[1], An T. D. Le[1], & Daniel T. Smith[2],

1. Centre for Brain Science, University of Essex.
2. Department of Psychology, University of Durham.

Running head: Spontaneous perspective taking?

Key words: Perspective taking – Vision – Theory of mind – Social attention

Submitted: 17/07/2015; Resubmitted: ##/10/2015

Total word count: 3829

Address for correspondence

Geoff G. Cole
Centre for Brain Science
University of Essex
Wivenhoe Park
Colchester
CO4 3 SQ
E-mail: ggcole@essex.ac.uk
Tel: 01206 872331
Fax: 01206 873801

Abstract

A growing number of authors have argued that humans automatically compute the mental state of other individuals. For instance, in the dot perspective task, observers are faster to judge the number of dots in a display when a human avatar has the same perspective as the observer compared to when their perspectives are different. This finding has been interpreted as evidence for 'spontaneous perspective taking'. The present examined this claim using a variant of the dot perspective paradigm in which we manipulated what the avatar could see via physical barriers that either allowed the targets to be seen by the avatar or occluded this view. We found a robust 'perspective taking' effect despite the avatar being unable to see the same stimuli as the participant. These findings do not support the notion that humans spontaneously take the perspective of others.

Introduction

As humans, we often make conscious judgments concerning the mental state of other individuals in social situations. This occurs, for instance, when one wonders why a person is gazing at a particular location. Furthermore, the computation of other people's perspective is central to efficient social cognition. A number of authors have argued that certain types of 'Theory of Mind' processes can occur automatically such that they are fast and do not require controlled processing. The most notable example was reported by Samson, Apperly, Braithwaite, Andrews, and Scott (2010), who argued that humans rapidly and spontaneously compute the perspective of other individuals. They employed a paradigm that has become known as the dot perspective task, in which observers are presented with a human avatar (located in the centre of a virtual room) that looks either towards a left or right-hand wall. A number of discs are positioned on the two lateral walls and the participant is asked to judge the number of discs from either their own perspective or the avatar's perspective. The central manipulation concerns the consistency of the avatar's and participant's perspective; on some trials the avatar and participant can see the same number of discs whilst on other trials they see a different number. For example, if the avatar looks to the right-hand wall and one disc is located on each of the two walls, the avatar sees one disc and the participant, by virtue of being able to see the whole room, sees two. By contrast, if two discs appear on the right-hand wall and none on the left, both the participant and the avatar see the same number of discs (i.e., two). Samson et al. found that reaction time (RT) to make the disc number judgment was shorter when the viewpoint of the avatar was consistent with the participant's relative to when their viewpoints were inconsistent. The authors concluded that this consistency effect occurs because the computation of another person's perspective occurs spontaneously.

In effect, the observer is said to compute what the avatar can see, and this representation includes the number of discs that can be seen. The knowledge about what the avatar sees then interferes with the observers' knowledge about the total number of discs present, thus increasing RT when the two are inconsistent.

The spontaneous visual perspective taking notion has not however gone unchallenged. For instance, Santiesteban, Catmur, Coughlan Hopkins, Bird, and Heyes (2014) argued that the avatars employed in the Samson et al. experiments act as a cue that shifts attention to one side of the display. Indeed, the basic spontaneous perspective taking method is similar to the classic central cueing paradigm (Langton & Bruce, 1998) in which a cue, for instance a human face, is located in the centre of a display and looks towards the left or right hand side. Furthermore, the critical comparison of the Samson et al. method, i.e., 'consistency-inconsistent' (of the avatar's and participant's viewpoint), maps directly onto the critical comparison in the central cueing paradigm, i.e., 'cued-uncued'. Although Samson et al. do include attentional cueing as a process that *contributes* to spontaneous perspective taking, a cueing effect could *solely* explain the basic effect. In support of their directional cueing hypothesis, Santiesteban et al. showed that a stimulus known to shift attention laterally (i.e., a centrally located arrow) induced consistency effects of comparable size to that of an avatar.

A problem however with the cueing hypothesis is that the perspective and cueing effects may operate independently but still generate a similar pattern of data. Thus, demonstrating that both arrows and avatars generate a consistency effect does not falsify the spontaneous perspective taking theory. As Firestone and Scholl (in press) have recently reminded us, "not only should you observe an effect when your theory calls for it, but you should also not observe an effect when your theory demands its absence". The principal aim of the present work was to test the claim that perspective taking is indeed spontaneous, as argued by Samson et al. (2010), by setting up a scenario in which visual perspective taking should not occur. As with the original experiment of Samson et al., participants were presented with an avatar located in the middle of a display that looked either towards a left of right hand wall. Importantly, the ability of the avatar to see the stimuli that generate the basic perspective taking effect was manipulated by the positioning of physical barriers either side of the avatar. On 'non-seeing' trials these barriers fully occluded stimuli presented to the left or right whilst, on 'seeing' trials the barriers included window-like features allowing the stimuli to be seen by the avatar (see Figure 1[1]). The use of physical barriers to manipulate what an agent can see is common when assessing mentalising in non-human animals (e.g., Hare, Call, & Tomasello, 2001). Clearly, if

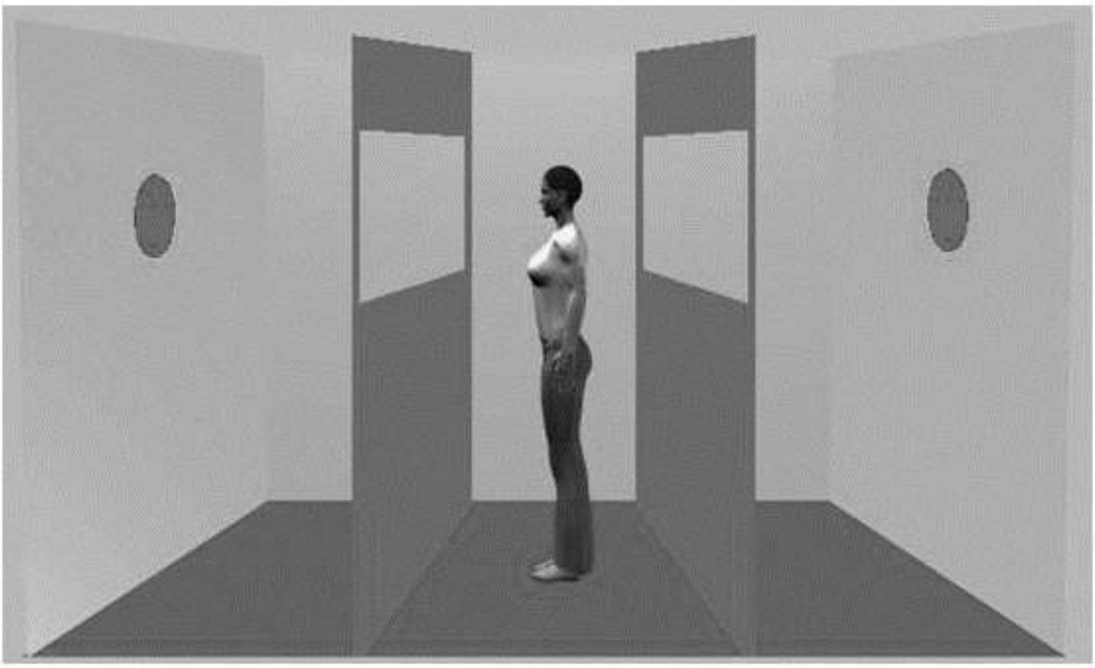

Figure 1. Stimuli used in the Experiment. The example shows a trial in which the avatar's view is inconsistent with that of the participant. We as the viewer can see the two discs but the avatar can only see one. The example also shows barriers in the 'seeing' condition; the avatar can see one of the walls. In 'non-seeing' trials the 'windows' of the barriers are closed.

the avatar's perspective is spontaneously taken, no Samson et al.-like effect should occur when the avatar is unable to see the inducing stimuli.

Method.

Participants. There were 24 participants who took part in exchange for course credit.

Stimuli and apparatus. The virtual room was 19.6° wide and 12° high. A male or female human avatar (7.8° in height) was located in the centre and always faced to the left or right-hand wall. Barriers were located to the left and right of the avatar and were approximately the same height as the room and were 1.8° wide. The barriers were solid on half the trials and thus prevented the avatar from seeing the wall being faced. On the remaining trials the barriers had a section cut out, allowing the wall to be visible. On the left and right-hand walls were a number of red discs (0, 1, 2, or 3) measuring approximately .7° in height. On 50% of trials, the avatar faced towards the same number of discs that the participant could see, whilst on the remaining trials the avatar faced towards a different number of discs. This manipulation is the same as Samson et al.'s (2010) 'consistent'-'inconsistent' manipulation. However, note that when the barriers occluded the avatar's lateral view, the participant's view was of course never consistent with the avatar's. We therefore consider this as manipulating whether the avatar faced towards the same or different number of discs as that of the participant. The room and barriers together with a black fixation cross were present as background throughout the entire experiment. As with Samson et al., male observers were presented with a male avatar and female observers were presented with a female avatar. The experiment was run on an Apple eMac computer linked to a CRT monitor.

Design and procedure. A within-participant, 2 x 2 factorial design was employed. The first factor manipulated whether the avatar faced towards the same number of discs that the participant could see or faced towards a different number ('same' vs 'different').

The second factor manipulated the avatar's vision of the room's left and right hand walls ('seeing' vs 'non-seeing'). Each trial began with the presentation of a number ('1', '2', or '3') located in the centre of the display for 750 ms. This informed the participant of the disc number in the display that needed to be verified on the current trial. For instance, when the number '2' appeared, this informed the participant that they will need to decide as quickly as possible whether two discs are present in the display. This number either matched the number of discs presented in that trial or did not match. This number disappeared for 500 ms after which the avatar and discs appeared until the participant responded. The beginning of a trial was initiated by the participant's response on the previous trial. The participant was asked to press a left-hand button if the disc number matched the number shown at the beginning of the trial or a right-hand button if they did not match. Observers were seated approximately 70 cms from the display and asked to respond as quickly as possible whilst keeping errors to a minimum. The visibility condition was blocked. Blocking this factor meant that attribution of what the barriers allowed the avatar to see did not need to occur trial-by-trial. At the beginning of each block, participants were shown an example of the relevant barrier and explicitly told that the avatar could either see or not see the two walls depending on which barrier/block was presented. There were 288 trials in total, 144 of which were match trials and 144 non-match trials. Half of the trials were 'same' and half were 'different'. For both matching and non-matching trials, there were 48 trials in which one disc was present, 48 trials in which two discs were present, and 48 trials in which three discs were present. We did not include what Samson et al. called 'filler' trials in which no discs were presented. Otherwise, our method closely replicates the aspects of Samson et al. critical to generating a spontaneous perspective taking effect. Twenty-four practice trials were given. Apart from the visibility condition which was blocked, and presentation order counterbalanced, all trial types were presented in a random order.

Check for the validity of our visibility manipulation.

Because our experiments were concerned with the central claim of Samson et al., i.e., rapid computation of what *others* see, we did not include trials used by Samson et al. in which participants were asked to take the perspective of the avatar. Indeed, *spontaneous* computation of others' perspective should not require observers to occasionally assume this perspective. This was alluded to recently by Schurz, Kronbichler, Weissengruber, Surtees, Samson, and Perner (2015) who argued that separating (i.e., 'blocking') trials in which participants are required to take their own perspective from trials in which they are required to take the avatar's perspective is more likely to index spontaneous perspective taking. Indeed, presenting both trial types within one block is likely to induce participants to explicitly (i.e., non-spontaneously) consider the avatar's perspective precisely because perspective taking is part of the task. Thus, in addition to our formal experiment we also ran a test to determine whether our visibility manipulation was effective. Five participants who did not take part in the main experiment were presented with 12 examples of the type of display shown in Figure 1. On six of these examples the window-like structures were open whilst on the other six they were closed. Participants were asked to take the perspective of the avatar and judge how many dots the avatar could see. All five participants were 100% correct. Thus, for instance, when the windows were closed all stated that the avatar could not see any dots.

Results

Three participant's data were excluded from further analysis due to an error rate of more than 20%. Our primary analysis was on RT; that is, the interval between avatar/dot onset and response. In line with Samson et al. (2010), only the matching trials were analysed. 4.2% of responses were outliers, defined as lying outside two standard deviations (SD) for each participant's condition mean, and omitted from further analysis. Figure 2 shows mean RTs for each of the four conditions. An ANOVA with number of discs faced with respect to the participant (same, different) and visibility (seeing or non-seeing) as within-participants factors revealed a significant main effect of discs faced, $F(1, 20) = 12.7$, $p < 0.002$, $\eta^2 = .39$, but no significant main effect of visibility, $F(1, 20) < 1$. The interaction was not significant, $F(1, 20) < 1$. We also analysed the error data using the same factors and levels described above. There was a small but non-significant main effect of discs faced, $F(1, 20) = 3.9$, $p < 0.06$, $\eta^2 = .2$, and no significant main effect of visibility, $F(1, 20) < 1$. The interaction was also non-significant, $F(1, 20) < 1$.

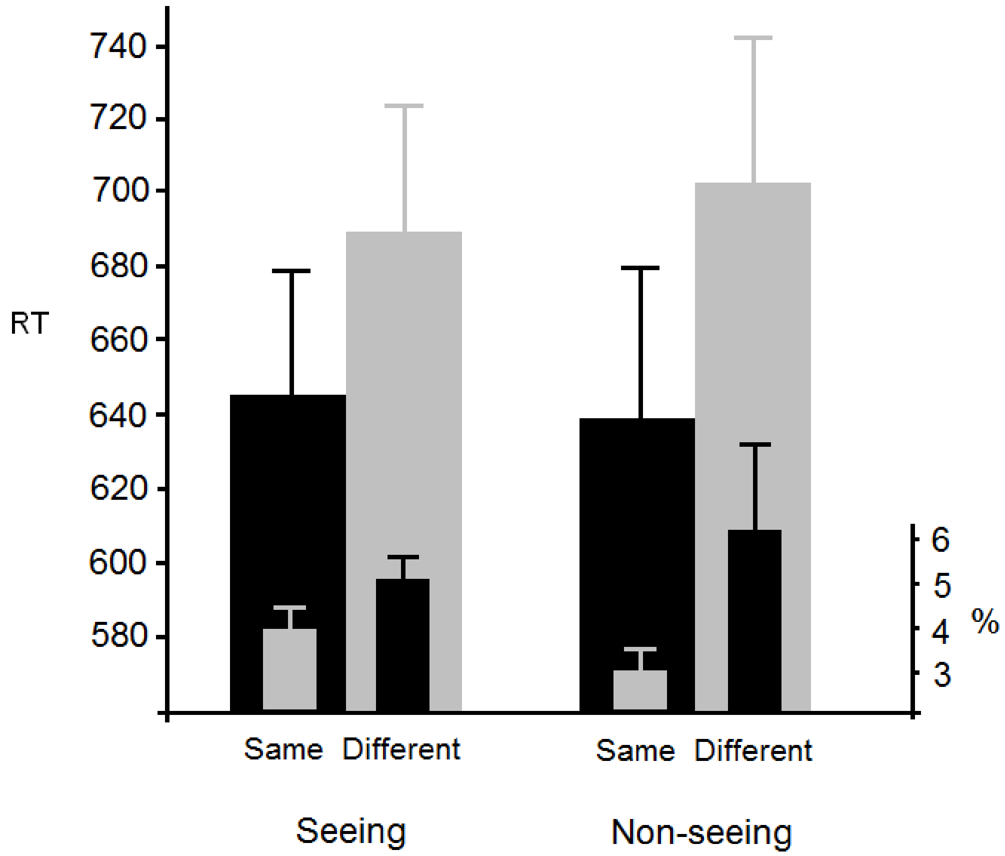

Figure 2. Mean RT and error rates together with standard errors.

Discussion

The present work examined whether observers spontaneously compute the perspective of other individuals, as reported by Samson et al. (2012). In addition to the standard dot perspective condition, we employed stimuli in which the avatar could not see any targets and thus never had the same perspective as the observer. As Samson et al. reported, we found that observers were relatively slow to make a perceptual judgement when their view of the critical stimuli was inconsistent with the view of another individual (i.e., an avatar) in the display. By contrast, RTs were relatively short when their view and the avatar's view were consistent. On their own, these data support the perspective taking account. Furthermore, the size of this effect is comparable to that reported previously (i.e., ~40 ms). However, data from trials in which the avatar could not see any of the two lateral walls, and hence never saw the same number of discs as the participant, also showed the same pattern of data. This suggests that the effect reported by Samson et al. cannot be due to the participant taking the avatar's perspective since no consistency effect should have occurred under this condition. This in turn challenges the spontaneous perspective taking account.

One particular advantage of the present study is that we have provided a very direct test of a spontaneous mental attribution theory. As we suggested in the Introduction, we have set up an experiment in which the spontaneous perspective taking theory can be falsified. These data are also consistent with our previous observation that changes in perspective do not modulate the gaze-cueing effect (Cole, Smith, & Atkinson, 2015). The results of our direct tests contrast with many previous studies that have *indirectly* assessed spontaneous ToM processes. This is particularly the case for neuroimaging studies. fMRI work suggest that the medial prefrontal cortex (mPFC) and the temporo-parietal junction (TPJ) are predominantly involved in social inferences of others, that is, thoughts, goals and intentions (e.g., Van Overwalle, 2009; Frith & Frith, 1999). Whilst measuring blood flow from a number of brain areas including the mPFC, Calder et al. (2002) presented observers with photographs of people whose eyes gazed at various positions. Participants were asked to indicate whether the models had thick or thin eyebrows. Importantly, observers were not asked to consider the mental state of the models; they were only asked to make a simple perceptual judgment. Results showed that direct gaze was particularly associated with activity in the fusiform gyrus. This can be expected given this region's well known involvement in face processing (Cole, Heywood, Kentridge, Fairholm, & Cowey, 2003; Kanwisher, Mcdermott, & Chun, 1997). More importantly however, was the observation that averted gaze lead to greater activity in the mPFC cortex. In other words, activity in a brain region associated with ToM was automatically activated when a person observed gaze behaviour. Although Calder et al. did suggest the possibility that participants may have attended to the models' gaze in order to explicitly consider their mental state, the authors did conclude that this inference was made automatically. Similarly, Schurz et al. (2015) have recently reported that participants undertaking the specific perspective taking paradigm of Samson et al. (2010) show pronounced activity in the mPFC and (right) TPJ. As with Calder et al. the authors argued that these result suggest spontaneous perspective taking. However, our current data suggest that this correlation between activity in certain brain areas and gaze direction cannot be taken as strong evidence for the automatic computation of perspective. Furthermore, and perhaps more significantly given the attention cueing explanation of the basic Samson et al. effect (Santiesteban et al. 2014), the TPJ has also been implicated in attention reorientation as well as self–other distinction (see Decety & Lamm, 2007; Mitchell, 2008).

Although the question is beyond the aims and scope of the present work (i.e., we have refuted the perspective taking claim rather than supported it), one outstanding question is why the consistency effect occurs at all. In addition to the cueing hypothesis (Santiesteban et al. 2014), another possibility concerns the critical stimuli (i.e., the dots) and their role as task relevant items. Perhaps observers *do* take on the perspective of the avatar, but this perspective only sees information relevant to the observer's task; it may be blind with respect to all other stimuli, including barriers. Samson et al. (2010) argued for the 'spontaneous' aspect of their account based on the fact that the avatar's perspective was task irrelevant. However, an abundance of work on *attentional control settings* in visual cognition has shown that such relevancy can operate in extremely subtle ways (Folk, Remington, & Jonhston, 1992). In the present context, having a strong top-down attentional set for discs could have resulted in participants assuming the avatar's perspective with respect to these stimuli. This set would then receive sensory reinforcement on every trial as participants made saccades to the dots but rarely, if ever, to the barriers. Although the authors were primarily concerned with *evolutionary* relevant stimuli, data from one previous spontaneous perspective taking study may provide support for the task relevancy hypothesis. Zwickel and Muller (2010) asked observers to rapidly localize a dot from either their own perspective or that of a face presented in the centre of the display. Results showed that RTs were slower when the adopted perspective and the observer's perspective conflicted. Importantly however was the observation that this only occurred when the agent exhibited a fearful expression. As with many other evolutionary important stimuli (e.g., venomous animals; see Cole & Wilkins, 2013), this type of emotive face is often considered as being of general relevancy.

Furthermore, task relevancy may account for why primates are sensitive to opaque barriers in visual perspective taking tasks but adults in the dot perspective task are not. In these studies, a subordinate animal will modify its behaviour based upon whether a dominant animal can see a mutual food reward. Again, the presence of evolutionarily relevant stimuli in the form of a more dominant agent may heighten the sensitivity of the subordinate ape to sources of occlusion in another's perspective (Hare, et al., 2001).

To summarize, the current study examined the claim that perspective taking occurs spontaneously. Contrary to this claim we observed a robust 'perspective taking effect' even when the perspective of the avatar was different to that of the observer. These results are not consistent with the claim that humans spontaneously compute perspective. It is worth noting that we do not argue that the dot perspective task does not involve any social processing mechanisms, only those concerned with spontaneous perspective taking. This is based on recent evidence by Nielsen, Slade, Levy, and Holmes (2015) who showed that the degree of spontaneous perspective taking positively correlates with self-reported empathy. Finally, future workers may want to consider employing the occluding barrier technique when using other paradigms suggestive of spontaneous perspective taking.

Footnotes

1). We are extremely grateful to Dana Samson for providing us with all her stimulus images, even though we only requested one example as a template to generate our own.

References

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A.W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40,* 1129–1138.

Cole, G. G., Heywood, C. A., Kentridge, R. W., Fairholm, I. & Cowey, A. (2003). Attentional capture by colour and motion in cerebral achromatopsia. *Neuropsychologia, 41,* 1837-1846.

Cole, G. G., & Wilkins, A. (2013). The fear of holes. *Psychological Science, 24*, 1980-1985.

Cole, G. G., Smith, D, & Atkinson, M. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception & Psychophysics, 77,* 1105-1115.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *The Neuroscientist, 13,* 580–593.

Firestone, C., & Scholl, B. J. (accepted target article). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences.*

Folk, C.L., Remington, R.W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030-1044.

Frith, C. D. & Frith, U. (1999). Interacting minds--a biological basis. *Science, 286,* 5445, 1692-1695.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61,* 139-151.

Kanwisher N., Mcdermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Langton, S., & Bruce, V. (1999). Reflexive orienting to social attention signals. *Visual Cognition, 6,* 541–568.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18,* 262–271.

Nielsen, M. K., Slade, L., Levy, J.P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Quarterly Journal of Experimental Psychology, 68,* 1931-1951.

Santiesteban, I., Catmur, C., Coughlan Hopkins, S., Bird, G., & Heyes, C. (2014). Avatars and Arrows: Implicit Mentalizing or Domain-General Processing? *Journal of Experimental Psychology: Human Perception and Performance, 40, 929-937.*

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1255-1266.

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D. & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage, 117,* 386-96.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30,* 829–858.

Zwickel, J. & Muller, H. J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition, 117,* 101-105.

END